SINGLE VALUED POLYGON MAPS

Martien Molenaar
Agricultural University Wageningen, Department of Surveying
Photogrammetry and Remote Sensing
P.O. Box 339
6700 AH Wageningen, The Netherlands
Commission Number: III/IV

Abstract.
Single valued rasters describe just one thematic aspect for a covered area. There is just one attribute value per position. This paper will give a definition of single valued polygon maps which has a great similarity with the concept of single valued rasters.
The information contained in a single valued polygon map can easily be structured according to the relational model and the network model for data bases. Information representation according to these models can serve as a first form of knowledge representation for map analysis. Therefore it is expected that the concept of single valued polygon maps can serve as a starting point for a branche of a geographical information theory for polygon data.

1. Introduction.

Several authors have dealt with the structuring of polygon data for geographical information systems and in their publications concepts and methods are given for this field of activities [2], [6], [7]. Proposals for database structures are based on these concepts and operational systems have been built [3], [4], [9]. Nevertheless a good theory structuring this field is still lacking or at least difficult to find from literature, so that a comparison of the different approaches is difficult and it is hardly possible to formulate a consistent set of criteria to evaluate the performance of database systems. Therefore this paper is meant to order the ideas found in literature into a conceptual base for a geographical information theory for polygon data.
The strategy for developing such a theory is to identify the elementary data types and the elementary links among them. The second step is then the definition of a query space being the set of all queries which can be answered by a system containing a database in which these elementary data types and links are represented.
When the query space is known it can be checked whether it contains the query set of a user, that is that all questions which a user would or could like to ask can be answered.
The third step is the interpretation of the elementary data types and links in the frame work of present database models. If such an interpretation is possible then the query language related to the database model can in principle cover the whole query space. Different database models can then be

compared with respect to their efficiency in handling the queries.
If not all data types and links can be represented a subspace of the query
space will be available for the user, it should be checked then whether this
subspace gives a better efficiency in handling certain queries.

2. Single valued rasters.

For storing geographic information two basic geometrical structures can be
used, e.g.: the raster structure and the polygon structure. The raster
structure is the more simple one. A two-dimensional grid of points or cells,
the raster elements, covers the terrain and each element has a set of
attributes to describe the thematic aspects at that position. Each attribute
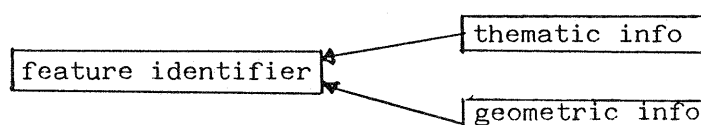represents a thematic aspect such as land use, soil type, population density
etc.

For the definition of basic operations in het processing of rasterdata, such
as data selection and overlaying, it is useful to start from the simplest
situation i.e. a raster in which the elements have only one attribute, that
is a "single valued raster". Depending on the scale type of the attribute
criteria can be formulated for data selection. In an overlay of single valued
rasters criteria for data selection depend on the combination of scape types.

These criteria for single valued rasters and overlays of single valued
rasters facilitate the formulation of a limited but complete set of elemen-
tary conditions for data selection. Complicated queries can be formulated
starting from these elementary conditions. Because of the fact that the set
of elementary conditions is limited it is easy to analyse the (potential)
information content of a database containing raster data. This means that
one can establish clearly which type of questions can be asked and will lead
to a clear answer and which questions will not lead to an answer.

3. Polygon data.

There is a basic difference between raster data and polygon data. The term
"polygon" will be used in a sense different from other publications, the
term refers in this text to geometric figure consisting of many sides which
meet pairwise under certain angles. In the previous section we saw that in
a raster the thematic information is stored per rasterelement and therefore
it is considered to be position dependent. A description by means of polygons
implies that besides thematic aspects of the terrain also geometric aspects
are taken into consideration. In rasters there is a direct link between
thematic information and position (geometry) for polygon data this link is
an indirect one in the sense that terrain features are identified and the
thematic information and the geometric information both are linked to
feature identifiers.

fig. 1.

feature identifier | thematic info
                     geometric info

As for raster data the structure of polygon data will be analysed for a two-dimensional situation. Three types of features can be distinguished in that case according to their dimensionality. These are the two-dimensional area features, the one-dimensional line features and the zero-dimensional point features.

The definition of single valued rasters prescribes how thematic information should be linked up with geometric (positional) information. Therefore the definition of single valued polygon "maps" should also prescribe how thematic information is linked up with geometric information. In this case the geometric information has three levels which should be discussed first.

Terrain features are described by polygons. For an area feature a polygon describes its boundary and for a line feature its linear structure. A point feature is simply represented by a point with its coordinates to define its position.
A polygon can be considered as a set of nodes, connected by edges, or alternatively as a set of edges which join at nodes. So at the first level the geometric information deals with the topology of the polygons, that is the connectivity of nodes and edges. The mathematical tool at this level is given by graph theory. The edges which build up a polygon do each have their own length and they meet under certain angles at the nodes. These lengths and angles give the geometric information at the second level, this level concerns the shape of the polygons. An alternative option is that the edges are not straight line segments as has been assumed here implicitly. They may have another shape which should then be described separately.
The third level concerns the positional information contained in the coordinates of the nodes.

It is clear that the edges and the nodes of the polygons are the basic elements of the geometric description of the features in a polygon map. So the definition of single valued polygon maps should describe how thematic information must be linked to these elements.

Each edge connects two nodes, the combination of a set of nodes and a set of edges forms a graph. In the sequel we will consider the edges as being directed, that means that it runs from a beginning node to an end node. In that case the term "arc" will be used instead of "edge". The graph is then a directed graph or a "digraph".
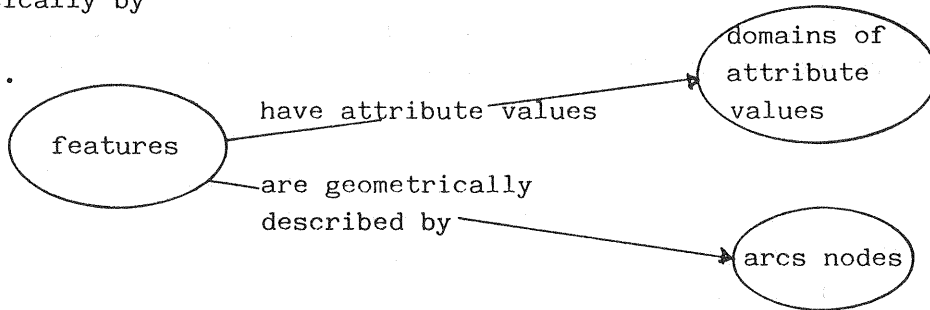
4. Single valued polygon maps.

The previous section shows that polygon maps contain three kinds of information, e.g.:
1. Feature identification.
2. Information concerning the thematic aspects of the features. The three feature types will each have their own attributes. For each type the attribute values may vary for the different features.

3. Geometric information given by the arcs and nodes constituting the poly-
   gons which describe the area- and line features, and the nodes represen-
   ting the point features.

This relationships among these three kinds of information can be given
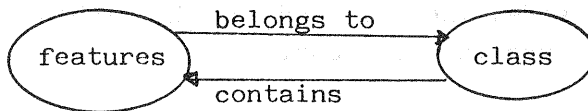scematically by

fig. 2.

The definition of a single valued raster allows only one attribute to be
linked to the elements of a raster. Hence at any position in the area
covered by the raster one finds only one attribute value.

In polygon maps the attributes belong to the features and through the
features they are related to the geometric elements. A first requirement for
a single valued polygon map should be that each feature type has only one
attribute and that the attribute values should be defined so that they are
mutually exclusive. The effect of this requirement is that each attribute
value defines a class of features and that these feature classes are dis-
joint. Each feature belongs to a class and each class contains zero or more
features. These relationships will be called "feature-class links", short
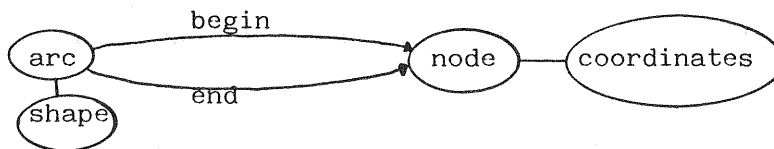"fc-links", they can be represented graphically by:

fig. 3.

The relationships among the geometric data will be called "geometry-
geometry links" or short "gg-links". There are four elementary gg-links:
- Each arc has a beginning node.
- Each arc has an end node.
- Each arc has a shape.
- Each node has a position (a pair of coordinates).
We will add the restriction that between each pair of nodes there is at most
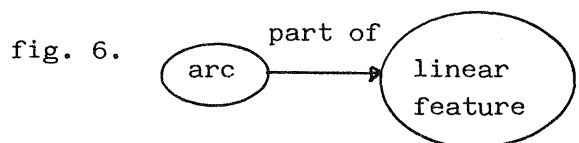one arc. A graphical presentation of these relations is given by:
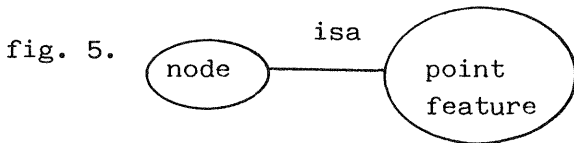
fig. 4.

From these elementary gg-links other relationships can be derived such as:
- Two arcs are connected = they have a node in common.
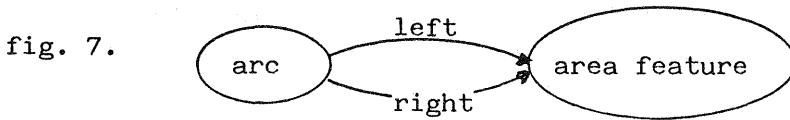- A polygon = a sequence of connected arcs.

- a polygon contains a node = a node is on a polygon = the polygon contains
  an arc which has the node as a beginning or an end node.

The relationships between the geometric elements and the features will be
called "geometry-feature links" or short "gf-links". They are given by the
fact that an area feature is geometrically represented by a polygon
describing its boundary, furtheron a line feature is represented by a
polygon describing its linear structure and a point feature is represented
by one node.

The graphical representation of the last gf-link is given in figure 5.
From the fact that a polygon is a sequence of connected arcs we find that an
arc may be a part of a polygon describing a linear feature, in graphical form
this gf-link is given in figure 6.

fig. 5.   node ——isa—— point feature   fig. 6.   arc ——part of—▶ linear feature

An arc can also be a part of a boundary polygon of an area feature which
implies that the arc must have an area feature at its right hand side and one
at its left hand side, in graphical form:

fig. 7.   arc ——left—▶ area feature
                ——right—▶

If an arc does not belong to a boundary polygon but to a linear feature which
intersects an area feature, then it still has an area feature at its lefthand
side and at its righthand side. The lefthand and the righthand gf-link both
refer to the same area feature in that case.

With this interpretation arcs are always linked to area features through the
"left" and "right" gf-links. They are not necessarily all related to linear
features though.
With these elementary gf-links we can formulate the following definition:
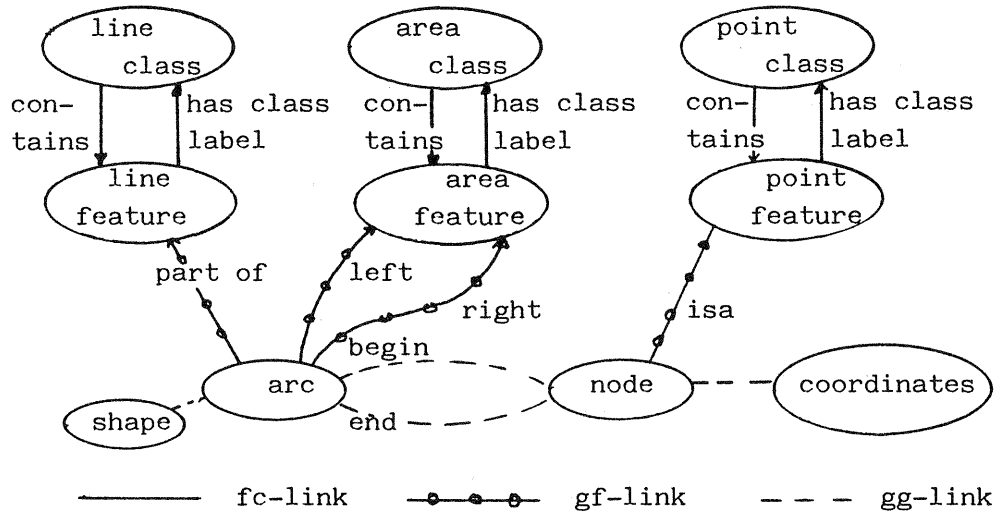
‖ In a single valued polygon map there is only one occurence of each
‖ elementary gf-link for each geometric element.

The consequence of this definition is that a node can only represent one
point feature and an arc can be part of at most one line feature and it has
just one area feature at its righthand side and just one area feature at its
lefthand side.

The link between a node and a point feature, may be an empty one, in the case
that the node does not represent any point feature. The same is true for the
link between an arc and a linear feature. To take care of such situations one
should use zero identifiers for point- and line features.

All the elementary link types identified in this section, are summarised in figure 8 in a graphical representation:

fig. 8.

This figure defines the information content of a database for a single valued polygon map. Now it is possible to analyse which queries can be applied to a database in which these data types and elementary link types are stored.

5. The query space of a single valued polygon map.

The total set of queries which can be handled by a database containing the elementary link types of the previous section we will call the "query space" spanned by the elementary link types. Each query in the space can be decomposed in the elementary link types given. Hereafter we will check for a number of queries whether they are contained in the query space.

The first group of queries use simple selection criteria, e.g.:

1> Give all area features of a specified class.
    The entry into the database is the area class, through the "contains" link the area features of this class are found and can be listed. If a graphical presentation is required then a plot file should be made. Therefore the relevant arcs and their shape are found through the "right" and "left" links. From the arcs the nodes and their coordinates found through the "begin" and "end" links.

Similarly line features and point features of a specified class can be found.

2> Give all features within a specified region.
    A region is specified by borderlines, then all nodes are selected within this region. Through the "isa" link the point features and their class labels within the region can be identified. Starting from the nodes the arcs connecting them are identified through the "begin" and "end" links. Then through the "part of" link the line features and their labels are

identified and through the "left" and "right" links the area features
and their labels are identified.

3> Give information about a specified feature.
   The entry is now a feature identifier. Through the "has class label" link
   the class label is found. The geometric information is found through the
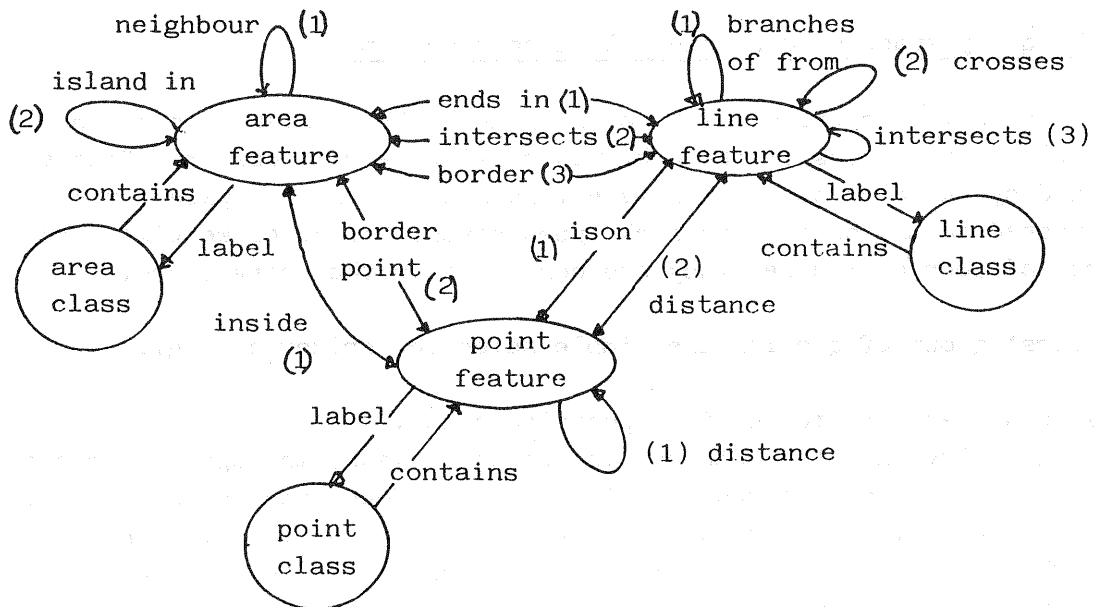   gf-links and the gg-links.

By combining the elementary commands of type 1>, 2> and 3> composite commands
can be constructed, such as:
>   Give all area features of a specified class in a specified region.
This command is a combination of 1> and 2>.

Many users of a geographic database are mainly interested in the information
at the semantic level, i.e.: the data concerning the features and the feature
classes. They may be interested in relationships among features. To structure
their queries we first have to identify the elementary relationships at this
level, then we can investigate whether these relationships can be decomposed
in the elementary links of the previous section.

The elementary relationships at semantic level are given in figure 9.

fig. 9.

Among the features there are six groups of relationships:

    area-area   →   aa-relationship = aa-r
    line-line   →   ll-relationship = ll-r
    point-point →   pp-relationship = pp-r
    area-line   →   al-relationship = al-r
    point-line   →   pl-relationship = pl-r
    point-area   →   pa-relationship = pa-r

- The al-r (2) and (3) can simply be decomposed in the "left" and "right"
  link and the "is part of" link of figure 8.

598

- The al-r (1) can be decomposed as follows: identify the arcs which are part of the line feature → "is part of", find the nodes of these arcs → "begin" and "end", find out which node occurs only once because that is an end node of the line feature, identify to which other arcs this node belongs → "begin" and "end", find out for which area features these arcs serve as a border → "left" and "right". Inspect the class labels of these area features, and check whether one of them is a candidate for the al-r (1) with the line feature. The search can also be done in the reverse way, starting from the area feature and finding the line features. In this way one can detect rivers ending in a lake and similar cases.
- The aa-r (1) and (2) can be analysed through the "left" and "right" links. If an area feature has only one neighbour it must be an island.
- The ll-r (1) and (3) can be decomposed in a combination of the "is part of" link and the "begin" and "end" links.
- The pl-r (1) can be decomposed in the "is part of" link and the "begin" and "end" links.
- The pa-r (2) can be decomposed in the "left" and "right" links and the "begin" and "end" links.
- The ll-r (2), the pl-r (2) and the pp-r (1) cannot be decomposed in the elementary links of figure 8. They can be derived from the information in the database through computational procedures making use of the coordinates of the nodes. This is also true for some feature characteristics such as the size of area features and the length of line features.
- The pa-r (1) can be handled in two ways. The relationship can be handeled through a computational procedure using the coordinates of the nodes and the shape information of the arcs involved. An alternative solution is that this relationship is added as a new link at feature level in figure 8. This new link is given in figure 10.

fig. 10.

feature-feature → ff-link

6. Single valued maps and database models.

The information analysis in the previous chapter resulted in a identification of elementary data types and the elementary links among them. The first aim of this exercise was to analyse the potential information content, i.e. the query space, of databases containing these data types and link types. A second task should be to interprete these results in the context of a data-base model. In this chapter two database models will be considered, i.e. the relational model and the network model.

For the relational model figures 8 and 10 should be interpreted in relations, which may represent entities and relationships. Several examples can be found in literature [7], [4], [9].
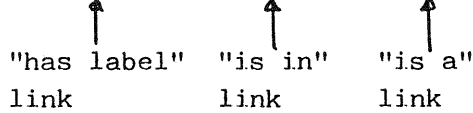We will give an interpretation in the notation (see [5] ch. 11):
    relation name (attribute 1, attribute 2, ...)

The full information content of figures 8 and 10 can be represented by the following relations (id. stands for identifier):
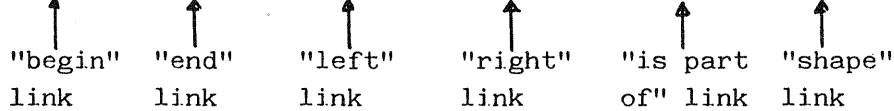
    area feature (area id., area class)
    line feature (line id., line class)
    point feature (point id., point class, area id., node nr)

                        "has label" "is in" "is a"
                        link   link   link

arc (node nr, node nr, area id., area id., line id., shape)

    "begin" "end"   "left"   "right"   "is part "shape"
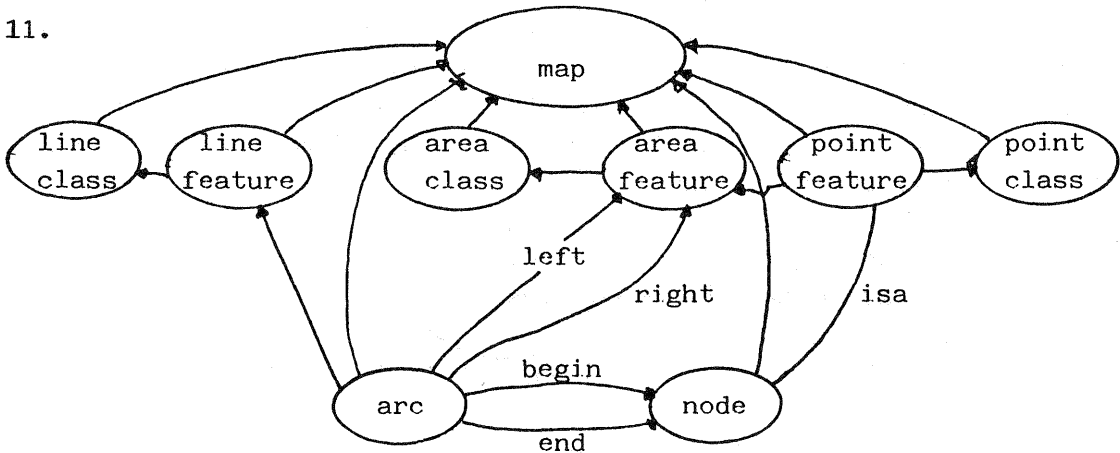   link   link   link   link   of" link link

node (node nr, x-coord., y-coord.)

The queries discussed in section 5 can now be applied to a database with this structure, using the operators given by Date (see [5] ch. 13).

For the network model figures 8 and 10 should be interpreted in record types and "many to one" links among these record types. This has been done in figure 11.

fig. 11.

In this figure the ellipses represent the record types and the arrows the many to one link types among them. These link types read as e.g.

 the map contains many arcs

 a class contains many features

 a line feature contains many arcs

 an area feature can be at the left hand side of many arcs

 a node can be the beginning node of many arcs

When the network database model uses pointer structures as given by Date
([5] ch 23.3), Ullman ([8] ch 3.2) then each record of a particular type
should have a field for a pointer for each link type in which this record
type participates.

From these two examples we see that the concept of single valued polygon
maps makes a straight forward definition of databases according to the
relational or network model possible. The query space for such databases can
simply be analysed. From section 5 it is clear that the potential information
content, may in fact exceed the query space defined in this paper, see e.g.
pp-r (1), ll-r (2) and pl-r (2). Further investigations should answer the
question whether such relationships should be treated separately or that the
query space can be extended so that they are contained in it.
In the definition we allowed only one attribute per feature, the class label.
If more information should be stored per feature then this could simply be
done by introducing an new link type in figures 8 and 10, that is a link
from a classname to an attribute list. In that case one could chose how many
attributes should be linked to each classname.

<div align="right">M. Molenaar, February 1988.</div>

References.

1. Broome, F.R.   : Mapping from a topologically encoded database 1.
                         The U.S. Bureau of census example.
                         Auto Carto London 1986.

2. Dangermond, J.   : A classification of software components commonly
                         used in geographic information systems.
                         Basic readings in GIS, spad systems ltd, 1984.

3. n.n.   : Technical description of the DIME System.
                         id.

4. n.n.   : ARC/INFO a modern geographic information system.
                         id.

5. Date, C.J.   : An introduction to Database Systems vol 1.
                         Addison-Wesley publ. comp. 1986.

6. Frank, A.   : Datenstrukturen für Landinformation systemen-
                         semantische, topologische und räumlicht Beziehungen
                         in Daten der Geo-Wissenschaften.
                         Inst. für Geodäsie und Photogrammetrie ETH Zürich
                         1983.

7. Roessel, J.W. van   : A relational approach to vector data structure
                         conversion.
                         Inst. Symposium on spatial data handling.
                         Zürich 1984.

8. Ullman, J.D.   : Principles of Data Base Systems.
                         Comp. Science Press 1982.

9. Waugh, T.C. and   : The Geoview design. A relational database approach
   R.G. Healy   to geographical data handling.
                         Int. Journal of Geographic information systems.
                         Vol 1. nr 2 pp 101-118.